Feature Selection Research for Electromyography (EMG) Classification

Clara Boothe Luce Summer Research Scholarship Final Report

Summer 2020

CBL Scholar: Declan Ketchum

Research Team: Rishita Sarin, Maya Sivanandan, Cory Knox

Supervisor: Dr. Samantha Michalka

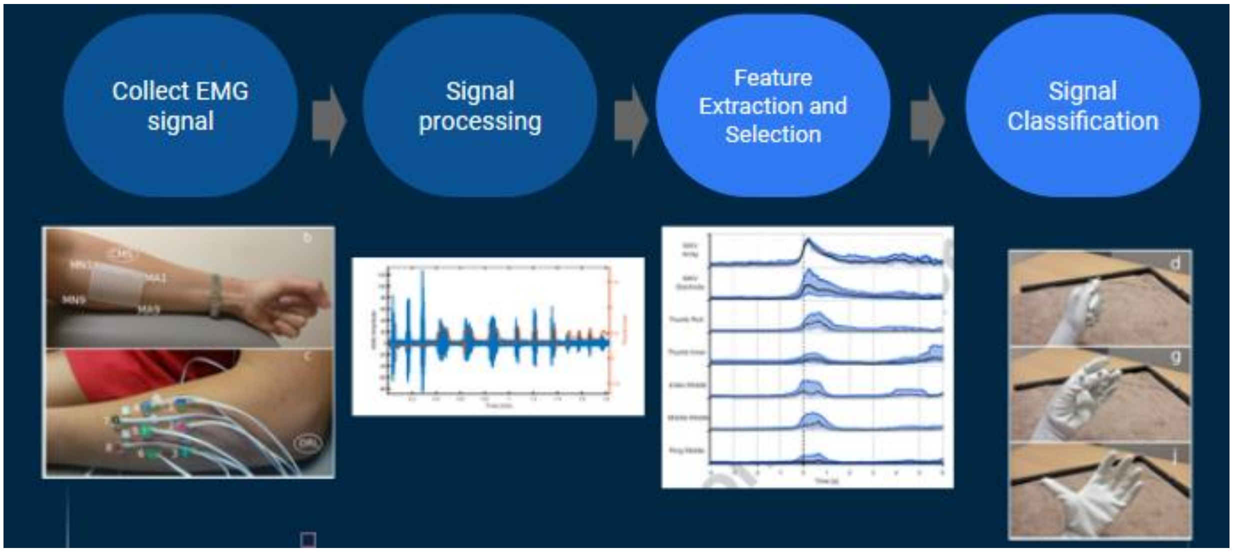
Abstract

Electromyography can be used as a human machine interface in which a person could control a computer or device with the electrical signals that cause muscle movement. This research used EMG data from the SEEDS data base to explore what features should be used by machine learning algorithms to accurately classify EMG data into which motion a subject is preforming. We extracted features from the EMG data and then ran three different feature selection algorithms to find which features were the most useful in classification. In the end, we found eight features that our various selection algorithms selected the most and concluded that those features would be good starting place when trying to classify EMG data.

Introduction

Electromyography (EMG) is the process of measuring the electrical signals sent from the brain to neurons in muscles that stimulate muscle movement. EMG is commonly used in the medical field to assess nerve damage in patients; it is a growing area of study in the technical field for its uses outside of medicine. These uses include wearable technology, gaming, and virtual reality. To use EMG signals as control systems for any of these areas the signals would need to be processed and classified accurately and in real time. This research project explores using machine learning on features from EMG signals to identify the hand position of a subject.

Electromyography usage can be broken down as depicted in the chart below. The signal is collected by placing electrodes on a subject's skin near neurons where the electrical activity is happening. The signal has to be processed to get the desired signal without electrical noise. Then features can be extracted from the signal. Features are pieces of information about a signal such as what is the maximum frequency or amplitude of a signal. These pieces of information are important to the next step where a machine learning algorithm classifies the EMG signal into which motion is happening.



Our work this summer has focused on calculating features and selecting for ones that seem to be useful in classifying the EMG signals. We used EMG data from the SEEDs [data set] which has EMG data collected from 25 subjects preforming 13 hand positions. We used this data to train our machine learning algorithm and test which features gave us the best classification rates.

Methods

The work I did this summer focused around making features selection algorithms. We extracted 32 features[i] from the raw EMG data, but we could not use all of them because that would cause the classification algorithm to overfit the data. Instead, we want to choose a small group of features that give the best classification rate of the EMG signals. We tried three different methods of feature selection and compared the features each chose for subjects 1-5 and what the classification accuracy each yielded.

The first selection method runs an SVM classification on each of the 32 features individually and select the eight features with the highest validation accuracy. It then runs the classification using the top eight features and returns an accuracy.

The second method runs the classification on all the features individually and then orders the features from highest to lowest classification accuracy. It runs through the ordered list of features adds each feature to a list on included features and runs the classification on the list of included features. It records the classification accuracy as each feature is added. The selected features include the first feature and the seven features that increase the classification accuracy the most when added.

The last algorithm starts with a list of all the features extracted from the data. It adds each feature, one at a time, to a list of included features and runs the classification algorithm on the included features. If the addition of a feature increases the classification accuracy, the algorithm keeps the feature in the included features list. If it does not affect the accuracy or makes the accuracy worse, it is removed from the list. At the end of this step, we have a list of features that passed the first selection step. Then we take that list and loop through it in reverse order. Each time we take one feature off the list and run the classification algorithm. If the accuracy returned is unaffected by removing the feature, that feature is removed from the included feature list. If the accuracy goes down because of the removed feature, we keep the feature on the included feature list. At the end, we have a shortened list of features that have proven helpful in classifying the EMG data. This method was created based on a feature selection method described in an EMG paper by Sara Abbaspour.

Results and discussion

We compared the three selection algorithms described above by testing each algorithm on all our extracted features and recording which features each algorithm selected and the classification accuracy those features yield for each subject. In this test we used the SVM classifier, both fast and slow data for each subject, and we included every sixth channel from the electrode grid and the final eight electrodes (1:6:126 127:134). Moving forward we should check that the features these algorithms selected are good for overall EMG classification by testing the features selected on one subject on a different subject’s data.

	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5
Best 8	64.53% dasdv rms mmav1 lscale stdv mav damv dvarv	74.36% mmav1 dasdv lscale damv rms stdv mav ld	68.38% mav mmav1 lscale rms ld meanfreq stdv damv	44.02% lscale mav ld rms mmav1 stdv var dasdv	67.95% dasdv lscale mav rms stdv mmav1 meanfreq mpv
Increasing 8	52.99% dasdv zeros wamp bp110t256 dvarv damv mmav1 meanfreq	75.64% mmav1 zeros dfa meanfreq mav bp110t256 mpv np	11.97% mav meanfreq mmav1 np medianfreq' stdv iemg dvarv	53.85% lscale zeros mav mpv rms meanfreq bp110t256 stdv	59.83% dasdv meanfreq var bp80t110 np lscale medianfreq bp40t56
Addition Subtraction	64.53% wamp zeros	77.35% meanfreq	67.52% np medianfreq	53.85% zeros var	61.54% dasdv lscale

	var mpv rms bp110t256 bp20t40	medianfreqva r mpv rms bp110t256 bp80t110	zeros var bp256t512 bp40t56 bp20t40	bp110t256 bp80t110 bp64t80 bp40t56 bp20t40 bp2t20	wamp Medianfreq zeros mpv mmav1 rms bp80t110 bp40t56 bp20t40 bp2t20

We extracted 32 features from the SEEDS EMG data. 25 of those features appear somewhere on the table above; meaning that 25 were chosen as good features to include by our selection algorithms. We think we should be using about 8 features to get the best classification accuracy for this data set size. This idea is supported by the addition subtraction algorithm which yielded between 7 and 12 features for each subject. To choose which 8 features to use going forward we found the 8 features that were selected the most by the selection algorithms. The table below shows how many times each feature was selected by our algorithm. **Moving forward we should use: mean absolute value (mav), modified mean absolute value (mmav1), root mean square (rms), L-scale (lscale), mean frequency (meanfreq), difference absolute standard deviation value (dasdv), standard deviation (stdv), and zeros.**[ii] Some of these features are similar to each other in how they are calculated so further work is needed to show whether that redundancy between frequently selected features affects the effectiveness of this set.

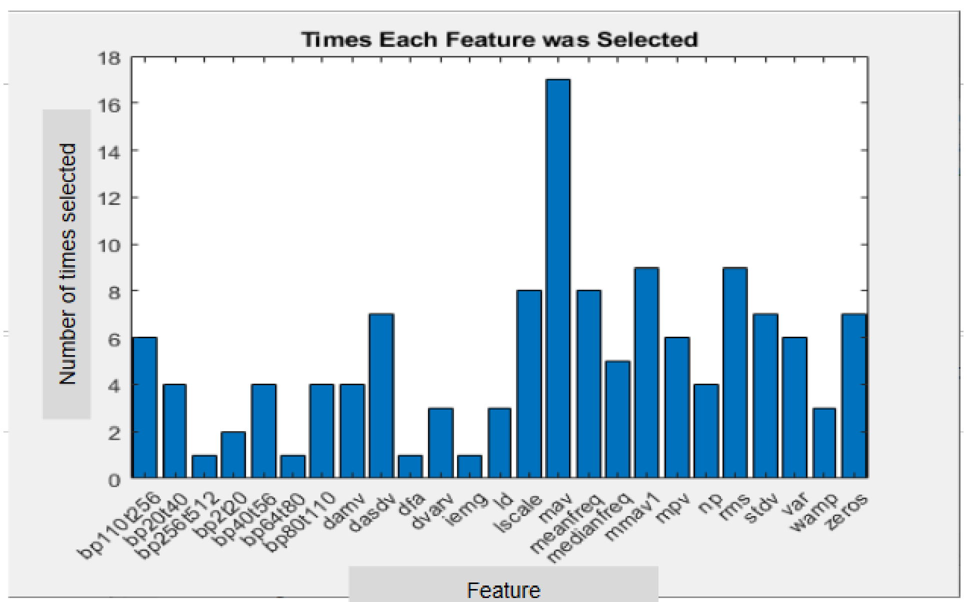


Figure 1: Bar graph with each figure and how many times our three selection algorithms selected that feature. We concluded that the 8 features that were selected the most were mav, mmav1, rms, lscale, meanfreq, dasdv, stdv, and zeros.

Acknowledgement

I was able to do this research because of the financial support of the Clare Boothe Luce Research Scholars Program at Olin College, funded by the Clare Boothe Luce Program of the Henry Luce Foundation. I received guidance from Dr. Samantha Michalka, and I learned from and with my research team: Rishita Sarin, Maya Sivanandan, and Cory Knox.

References:

Matran-Fernandez, Ana, et al. "SEEDS, Simultaneous Recordings of High-Density EMG and Finger Joint Angles during Multiple Hand Movements." *Scientific Data*, vol. 6, no. 1, 2019, doi:10.1038/s41597-019-0200-9.

Abbaspour, Sara, et al. "Evaluation of Surface EMG-Based Recognition Algorithms for Decoding Hand Movements." *Medical & Biological Engineering & Computing*, vol. 58, no. 1, 2019, pp. 83–100., doi:10.1007/s11517-019-02073-z.

---

[i] To see what these 32 features are, where we got them, and how we implemented them please look to this spreadsheet of information: https://docs.google.com/spreadsheets/d/1djwtkc51Tmc68BEfznd0i4aupOzRQDOBnz3kEGgRyKg/edit?usp=sharing

[ii] To see what these 32 features are, where we got them, and how we implemented them please look to this spreadsheet of information: https://docs.google.com/spreadsheets/d/1djwtkc51Tmc68BEfznd0i4aupOzRQDOBnz3kEGgRyKg/edit?usp=sharing